

L'evoluzione del ruolo delle Biblioteche Digitali Scientifiche (BDS)

Costantino Thanos

Istituto di Scienza e Tecnologie dell'Informazione "Alessandro Faedo"

La Rete di Eccellenza DELOS (2004 – 2008), finanziata dal Sesto Programma Quadro (6FP) della Commissione Europea, nell'arco della sua vita ha sviluppato due definizioni/visioni per la Biblioteca Digitale Scientifica (BDS):

Prima definizione/visione (San Cassiano Workshop 2001):

«Le Biblioteche Digitali dovrebbero consentire a qualsiasi cittadino di accedere alla conoscenza umana in qualunque momento e da qualunque luogo in un modo amichevole, multimodale, efficiente ed efficace superando barriere linguistiche, culturali, e di distanza ed utilizzando dispositivi interconnessi all'Internet».

Da notare che in questa Visione l'enfasi viene posta sull'"accesso" che a quel tempo era il principale obiettivo dell'internet e del web.

Seconda definizione/visione (Corvara, Workshop 2004):

«Le Biblioteche Digitali hanno il potenziale di diventare i depositi della conoscenza universale e i condotti di comunicazione per il futuro, un normale veicolo attraverso il quale tutti potranno accedere, discutere e accrescere/migliorare informazioni di tutte le forme».

Questa seconda visione tiene conto della rapida evoluzione dell'internet e del web da piattaforma di pubblicazione, dove l'aspetto im-

portante era l'accesso all'informazione, a piattaforma di "social media", dove l'aspetto importante è la creazione e lo scambio d'informazioni e la collaborazione.

In tutte e due le definizioni/visioni era implicito che il record scientifico era composto, fondamentalmente, dalla letteratura scientifica, dalla letteratura grigia e dai giornali scientifici. Essenzialmente, il record scientifico era di natura testuale.

Negli ultimi anni, significativi progressi tecnologici hanno modificato radicalmente la composizione del record scientifico. Alludiamo a progressi tecnologici che hanno permesso la raccolta (attraverso osservazioni), la produzione (attraverso esperimenti), e il calcolo (attraverso simulazioni) di enormi volumi di dati scientifici noti come big data. Altri progressi tecnologici hanno permesso la digitalizzazione di questi dati su larga scala. Di conseguenza, il record scientifico moderno è composto, fondamentalmente, da un gran numero di collezioni di dati scientifici (datasets) distribuite su scala globale e contenenti enormi volumi di dati dell'ordine di terabytes. Dati che presentano alcune importanti caratteristiche, per esempio dinamicità, grande dimensionalità e connettività. Queste collezioni di dati sono poco connesse fra loro e, di conseguenza, anche il record scientifico risulta poco connesso.

Le conseguenze di questa radicale modifica del record scientifico sono due, una di natura culturale e l'altra di natura tecnologica. La prima, di natura culturale, riguarda l'emergere

di un nuovo paradigma del metodo scientifico dovuto al fatto che gran parte della conoscenza scientifica prodotta dai ricercatori si è trasferita dalla letteratura ai dati. Di conseguenza, è nei dati che va cercata la conoscenza scientifica prodotta da osservazioni/esperimenti/simulazioni condotte dai ricercatori. Il nuovo paradigma, noto come data science, è basato sull'intenso utilizzo dei dati scientifici. La seconda conseguenza, di natura tecnologica, riguarda l'inadeguatezza di alcune tecnologie tradizionali riguardanti, principalmente, la ricerca (query processing) e gestione dei dati (data management) nel nuovo contesto tecnologico creato dai big data. Questa inadeguatezza ha motivato lo sviluppo di nuove tecnologie nella ricerca dei dati (data seeking) e nella loro interconnessione (data linking). Inoltre, nuovi approcci alla creazione di conoscenza stanno emergendo basati su tecnologie che supportano l'esplorazione di spazi dati interconnessi.

Infatti, il nuovo paradigma di metodo scientifico imporrebbe ai ricercatori/studiosi di esplorare il record scientifico creando percorsi di conoscenza (knowledge patterns) che permetterebbero loro di acquisire nuove conoscenze e/o formulare nuove ipotesi circa un fenomeno sotto studio. Questa esigenza, però, si scontra con il fatto che il record scientifico è poco connesso.

In questo emergente contesto culturale-tecnologico il ruolo della BDS è destinato ad un radicale cambiamento. In sostanza, la BDS non è più soltanto un deposito della conoscenza scientifica contenuta nella letteratura, ma viene ad assumere anche la funzione di una infrastruttura che supporta i ricercatori/studiosi nelle loro attività richieste dal nuovo paradigma di metodo scientifico. Quindi, la BDS deve supportare i ricercatori/studiosi nella loro attività che mira a scoprire e ad accedere alla conoscenza confinata nei datasets che si custodiscono in depositi disciplinari (domain-specific data centers) distribuiti su scala globale. La

BDS deve fornire servizi che permettono ai ricercatori/studiosi di scoprire esistenti relazioni fra dataset globalmente distribuiti e stabilire legami (links) fra questi creando così grafi di conoscenza, dove i nodi dei grafi sono dataset e gli archi rappresentano relazioni (semantiche/temporali/spaziali/causali) fra loro. La moderna BDS, in sintesi, deve fornire i strumenti che permettano ad un ricercatore/studiose sia di creare percorsi di conoscenza, sia di navigare in uno spazio dati interconnesso.

Una funzione molto importante, che in questo suo nuovo ruolo la BDS deve fornire, è quella della Pubblicazione dei Dati. Per pubblicazione dei dati si intende un processo che ha come obiettivo quello di permettere ai ricercatori/studiosi di scoprire, accedere, capire, riutilizzare e riprodurre i dati scientifici contenuti nelle varie collezioni dati distribuite su scala globale. Fasi importanti di questo processo sono la registrazione dei dati ed il loro potenziamento semantico.

Il nuovo ruolo della BDS, sopra tratteggiato, è destinato nel prossimo futuro a diventare prevalente rispetto al suo ruolo tradizionale, quello che la considera come un deposito di letteratura scientifica.

Un'ultima fondamentale considerazione riguarda il contesto politico all'interno del quale il nuovo ruolo della BDS sta emergendo. Questo contesto consiste nella politica di Scienza Aperta (Open Science). Per Scienza Aperta si intende una politica che prescrive la condivisione aperta di tutti i risultati della ricerca al più presto possibile compatibilmente con i tempi del processo della indagine/ricerca.

Tale politica è stata promulgata dalla Commissione dell'Unione Europea ed è diventata la politica che regola tutte le attività di ricerca condotte con finanziamenti Europei.

È evidente che il nuovo ruolo della BDS, sopra delineato, definisce anche un nuovo settore di ricerca. Infatti, affinché una BDS possa con efficacia svolgere il suo nuovo ruolo è necessario lo sviluppo di importanti attività di ricer-

ca. Qui di seguito elenchiamo una lista, ovviamente non esaustiva, di attività di ricerca necessarie, al nostro avviso, per lo sviluppo della BDS della prossima generazione.

- Per la creazione di un record scientifico interconnesso è fondamentale una ricerca avente come obiettivo l'allargamento del numero di tipi di relazione fra dataset (semantica, causale, temporale, spaziale ecc.) e di "assiomatizzarle" mediante anche nuove ontologie, specifiche per ogni disciplina scientifica. È, quindi, necessario anche lo sviluppo di motori di inferenza in grado di inferire relazioni di vario tipo fra dataset.
- Indirizzare la ricerca su tematiche riguardanti la citazione ed il collegamento di dati scientifici dinamici, in quanto una importante caratteristica dei big data è la loro dinamicità. Una tecnologia di citazione dinamica per i dataset scientifici è di fondamentale importanza per l'accesso alla conoscenza che si trova confinata nei big data.
- Sviluppare nuovi paradigmi di ricerca di dati (data seeking) e di creazione di cono-

scenza (knowledge creation) basati su approcci esplorativi che sfruttano anche la tecnologia di *machine learning* in quando i tradizionali approcci (query processing e data mining) sono inadeguati nel contesto dei big data.

- Indirizzare la ricerca verso la definizione di metadati, specifici per ogni disciplina, semanticamente potenziati.
- Indirizzare la ricerca su tematiche riguardanti la gestione (management) e attraversamento di grafi (graph data).
- Sviluppare ricerche verso tecniche di *data assimilation* e *correlation*.
- Sviluppare nuovi modelli di pubblicazione.

Ovviamente, come già accennato, questa non è una lista esaustiva ed in particolare gli aspetti multimediali dei big data meritano attento studio.

Una Conferenza sulle Biblioteche Digitali Scientifiche può costituire un forum dove queste tematiche vengono dibattute e dove relativi risultati vengono presentati.